ANP-D0449

DATA ANALYSIS USING PYTHON

TWITTER HASHTAG ANALYSIS: UNDERSTANDING ENGAGEMENT TRENDS

SUBMITTED BY:

MEENA P

ABSTRACT:

Twitter is a dynamic social media platform where hashtags play a crucial role in content discovery and trend analysis. Understanding engagement patterns associated with hashtags is essential for businesses, influencers, and researchers to optimize their strategies. However, the vast amount of daily-generated data makes it challenging to track and analyze trends effectively. This study focuses on Twitter hashtag analysis using text processing, frequency distribution analysis, data aggregation, visualization, and trend detection. By extracting hashtags from tweets, analyzing their frequency, and aggregating engagement metrics, we identify the most influential and viral hashtags. Visualization techniques, such as bar charts, enhance interpretability, while trend detection methods help differentiate between short-lived viral trends and long-term engagement patterns. This approach enables data-driven decision-making in social media marketing, content creation, and audience engagement. The insights gained from this study can support predictive analytics for future trend forecasting and improved social media outreach.

Problem Statement

- Twitter generates vast amounts of data with hashtags playing a key role in content categorization and visibility.
- Understanding the impact of hashtags on engagement metrics like likes and retweets remains a challenge.
- Businesses, marketers, and analysts struggle to identify the most effective hashtags for their strategies.
- There is a need for systematic analysis to determine which hashtags drive the highest engagement.
- This project aims to analyze hashtag usage patterns, measure their influence on engagement, and provide data-driven insights to enhance content reach and interaction.

Solution Approach

Data Collection:

- Use a dataset containing tweets, likes, and retweets for analysis.
- Load the dataset from a CSV file for processing.

Hashtag Extraction:

- Use regular expressions to extract hashtags from tweets.
- Store extracted hashtags for further analysis.

Frequency Analysis:

- Count occurrences of each hashtag in the dataset.
- Identify the most frequently used hashtags.

Engagement Analysis:

- Aggregate likes and retweets for each hashtag.
- Determine which hashtags drive the highest engagement.

Data Visualization:

- Use bar charts to visualize:

- Most frequently used hashtags.
- Hashtags with the highest likes.
- Hashtags with the most retweets.

Trend Detection & Insights:

- o Identify emerging or viral hashtags based on sudden spikes in usage.
- o Provide insights to marketers and analysts to improve content strategy.

Implementation:

```python
import pandas as pd

import re

import matplotlib.pyplot as plt

from collections import Counter

# Load dataset

df = pd.read_csv("twitter_large_dataset.csv")


# Function to extract hashtags

def extract_hashtags(tweet):

    return re.findall(r'#\w+', tweet)


df['Hashtags'] = df['Tweet'].apply(extract_hashtags)


# Flatten hashtag list

all_hashtags = [hashtag for hashtags in df['Hashtags'] for hashtag in hashtags]


# Frequency distribution
```

```python
hashtag_counts = Counter(all_hashtags)

hashtag_df = pd.DataFrame(hashtag_counts.items(), columns=['Hashtag',
'Count'])


# Aggregate engagement per hashtag

engagement_data = []

for index, row in df.iterrows():

    for hashtag in row['Hashtags']:

        engagement_data.append({'Hashtag': hashtag, 'Likes': row['Likes'],
'Retweets': row['Retweets']})


engagement_df = pd.DataFrame(engagement_data)

aggr_engagement = engagement_df.groupby('Hashtag').sum().reset_index()


# Visualization

plt.figure(figsize=(10, 5))

hashtag_df.sort_values(by='Count', ascending=False).head(10).plot(

    x='Hashtag', y='Count', kind='bar', legend=False, color='skyblue',
ax=plt.gca()

)

plt.title("Top Hashtags by Frequency")

plt.ylabel("Count")

plt.xticks(rotation=45)

plt.show()
```

```python
plt.figure(figsize=(10, 5))

aggr_engagement.sort_values(by='Likes', ascending=False).head(10).plot(
    x='Hashtag', y='Likes', kind='bar', legend=False, color='lightcoral',
ax=plt.gca()
)

plt.title("Top Hashtags by Likes")

plt.ylabel("Likes")

plt.xticks(rotation=45)

plt.show()


plt.figure(figsize=(10, 5))

aggr_engagement.sort_values(by='Retweets', ascending=False).head(10).plot(
    x='Hashtag', y='Retweets', kind='bar', legend=False, color='seagreen',
ax=plt.gca()
)

plt.title("Top Hashtags by Retweets")

plt.ylabel("Retweets")

plt.xticks(rotation=45)

plt.show()
```
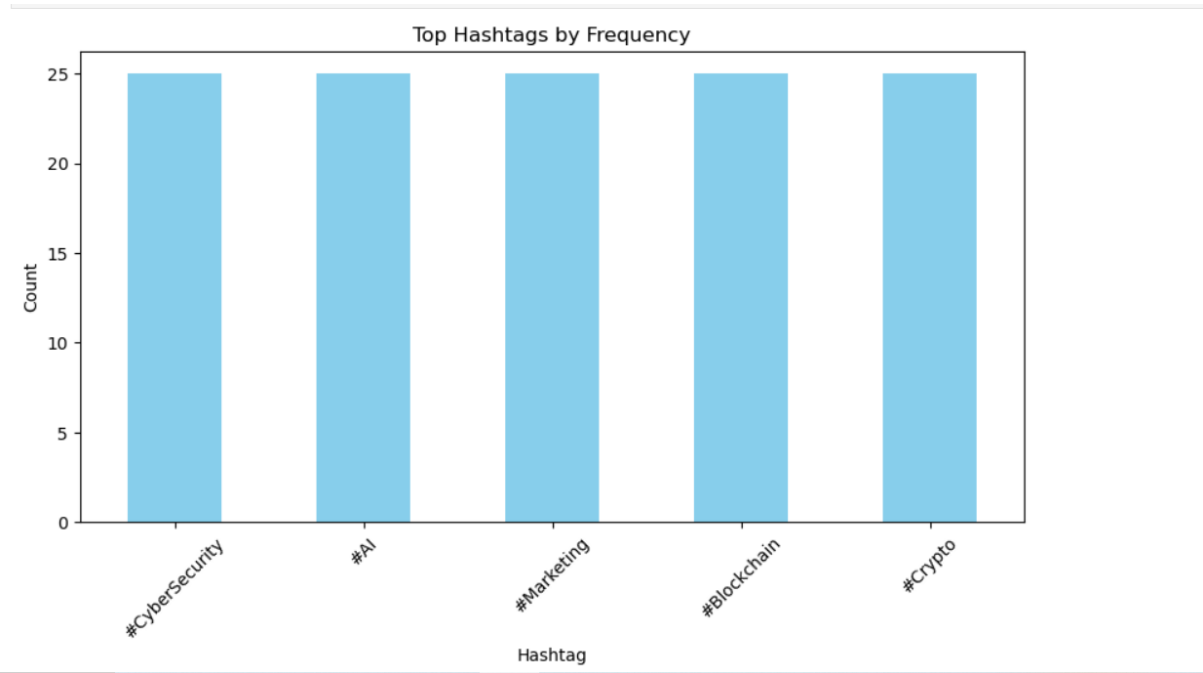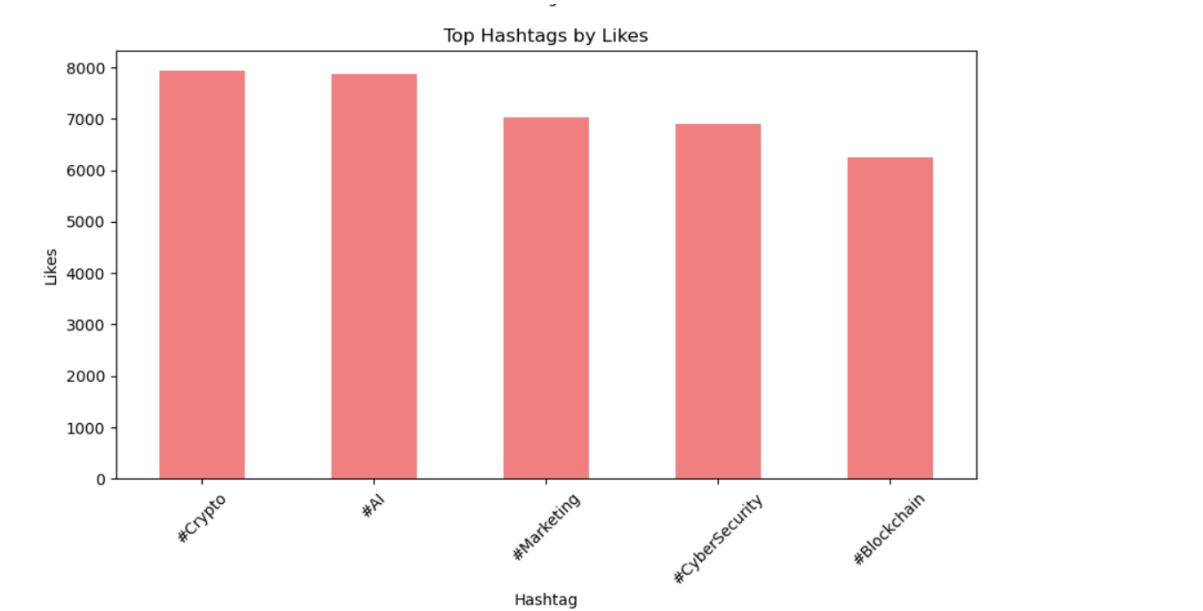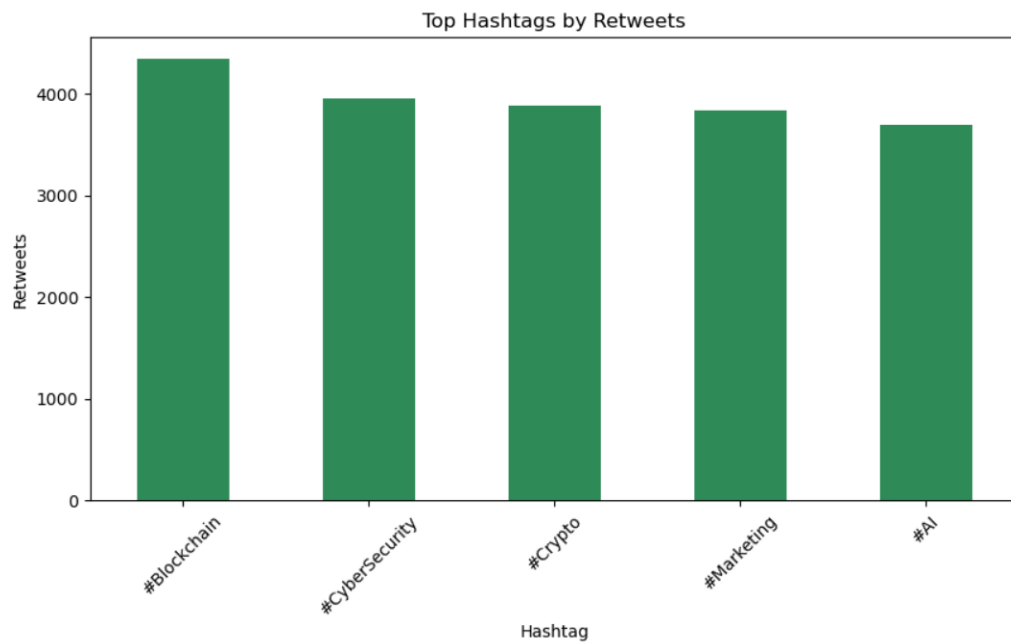
Output Visualization:

Top Hashtags by Frequency

Top Hashtags by Likes

Top Hashtags by Retweets

Top Hashtags by Retweets

Conclusion:

Analyzing Twitter hashtags provides valuable insights into engagement patterns, helping businesses, marketers, and researchers enhance their content strategies. By identifying the most frequently used and high-engagement hashtags, users can optimize their reach and visibility on social media platforms. This study demonstrates the significance of data-driven analysis in understanding social media trends and improving online presence. The findings highlight the importance of hashtag performance metrics in shaping effective digital marketing strategies. Future research can focus on real-time trend prediction and sentiment analysis to gain deeper insights into audience behavior and engagement.